

Diagramas Causales

Establece tus supuestos antes de tus conclusiones



Reynier Rodríguez Socarrás

reynier.socarras@uadec.edu.mx

¿Qué es la causalidad?

¿Qué es una relación causal?

Relación entre un evento (**la causa**) y un segundo evento (**el efecto**) donde el **efecto** es una consecuencia **directa** de la **causa**.

Estamos inmersos en **causa** y **efecto**.

- Si **riego las plantas** (**causa**), entonces **crecerán sanas** (**efecto**).
- Si **estudio toda la noche** (**causa**), **tendré una buena nota** (**efecto**).
- Si **rompo un espejo** (**causa**) **tendré siete años de mala suerte** (**efecto**).

La **causalidad** nos permite razonar sobre el mundo y desempeña un papel fundamental en la **toma de decisiones**.

¿Qué es la causalidad?

¿Por qué es importante la causalidad?

- No es suficiente el poder **predecir** la ocurrencia de un evento para **entenderlo**.
- Comprender la **estructura causal** de un proceso es un ingrediente necesario para el **conocimiento** sobre este proceso.
- Al comprender la causalidad podemos diseñar **intervenciones** para cambiar los **resultados**.
- Ser capaz de responder preguntas del tipo **por qué** es la esencia de la explicación científica.

Juega un papel esencial en problemas relevantes de investigación:

- ¿Ir a una escuela cara me asegura tener **éxito en la vida**?
- ¿La inmigración reduce mis posibilidades de **conseguir un trabajo**?
- ¿Las becas monetarias a los pobres reducen la **tasa de delincuencia**?

Inferencia Estadística vs Inferencia Causal

Inferencia Estadística

Muestra de datos → Población

- Obtenemos resultados para una **muestra** que se pueden generalizar para la **toda la población**, bajo un cierto **error**.
- El problema de la **inferencia estadística** va desapareciendo a medida que se incorporan más **datos**.

Inferencia Causal

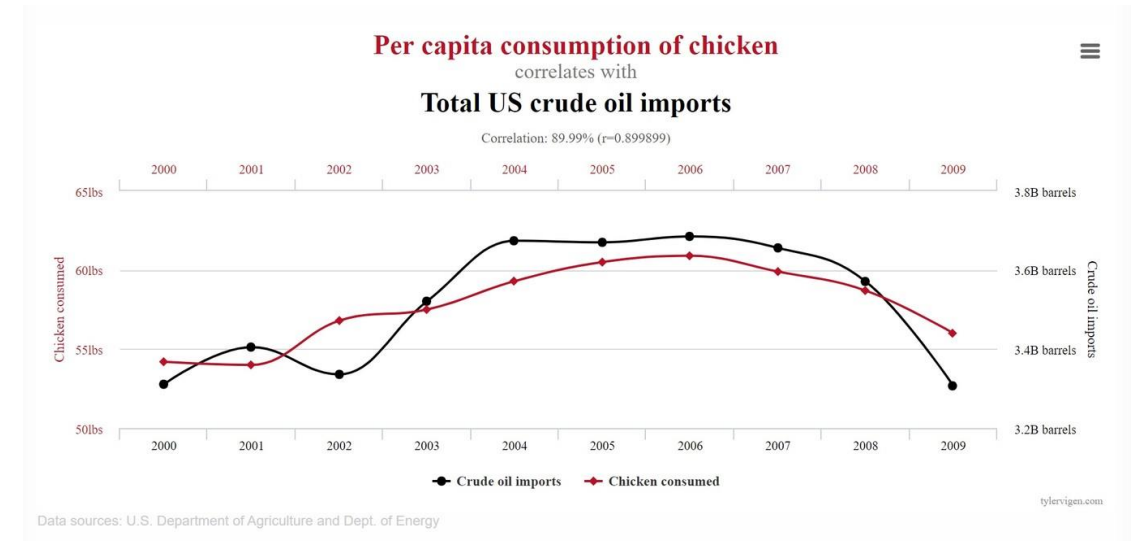
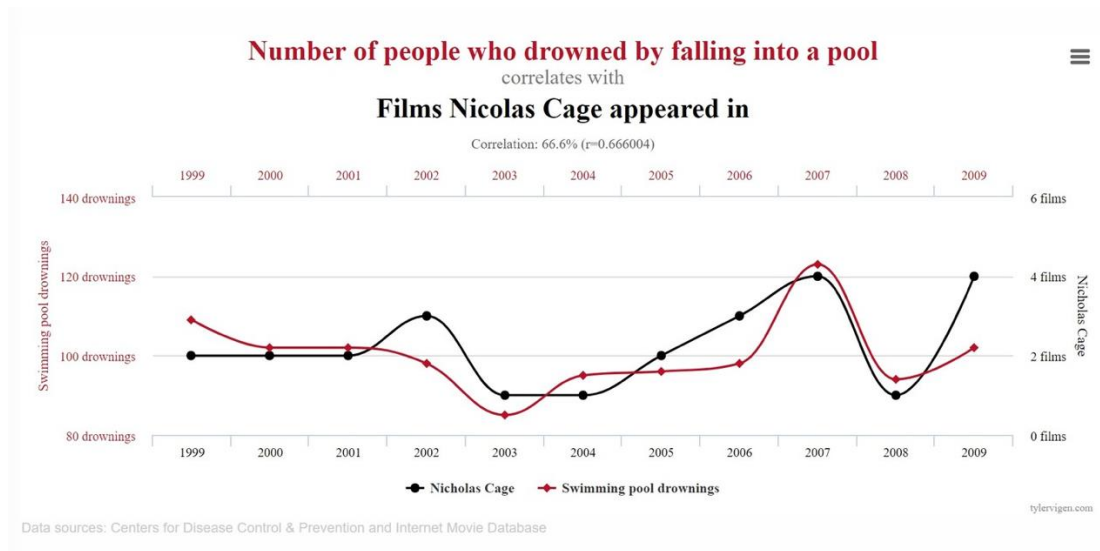
- Incluso suponiendo una base de datos de **toda la población** y estando seguros que existe una **relación** entre el **evento A** y el **evento B**.
- No podemos asegurar que **A** tiene un **efecto** sobre **B**.
- Para saber si existe una **relación causal** debemos conocer el **contrafactual**.

Problema fundamental de la inferencia causal

Contrafactual

- Que hubiese pasado con el **evento B** si el **evento A** nunca hubiese sucedido.
 - ¿Qué hubiera pasado con **mi perspectiva laboral** si hubiera elegido otra **carrera universitaria**?
 - ¿Cuánto habría crecido el **PIB** de México si no hubiera **firmado el Tratado de Libre Comercio**?
 - ¿Los **programas de renta básica universal** reducen la **desigualdad económica** en comunidades rurales?
- Para dos tratamientos diferentes **nunca** podemos observar **ambos resultados** de manera simultánea, ni la diferencia entre ellos.
 - En las comunidades donde se aplique el programa, nunca se podrá conocer que **hubiese pasado** si no se hubiese aplicado.
 - Existe un resultado observado (**factual**) y uno no observado (**contrafactual**).

¿Cómo establecer causalidad?



Correlación (asociación) no implica causalidad

- La **asociación** es una **precondición** para la causalidad.
 - Pueden existir relaciones **no lineales** (lo cual no incluiría a la correlación).
Causalidad \implies Asociación
Causalidad \nleftarrow Asociación
- La asociación entre variables puede darse deberse al azar.

¿Cómo establecer causalidad?

Consumo de helados

Golpes de calor



Tomar café



Enfermedades cardiovasculares

Asociaciones espurias

- Una asociación (correlación) **espuria** es una relación entre dos variables que se mantiene incluso cuando la muestra es grande, pero que en realidad **no** existe.
- Sesgo sistemático

La temporalidad en la causalidad

- Cantidad de personas usando paraguas cada día.
- Condiciones meteorológicas para cada día.
- En los días lluviosos hay más personas usando paraguas.
- ¿Los paraguas provocan la lluvia?

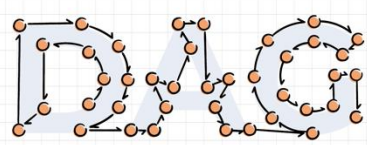


Tláloc



- La **precedencia temporal** es otra precondition para la causalidad
- Mientras la asociación (correlación) es **simétrica**, la causalidad **no** lo es
- Existe un tipo de diagrama que cumple con estas características y nos ayuda a representar las relaciones causales.

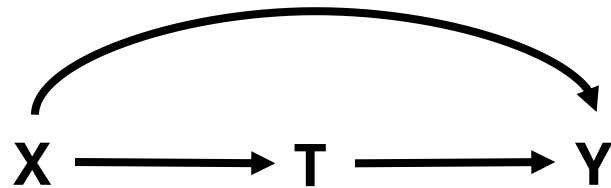
Directed Acyclic Graph



Grafos Acíclicos Dirigidos

Diagramas usados para el modelamiento de relaciones de causalidad.

- Los nodos representan variables.
- Las flechas o aristas indican relaciones causales.
- Caminos (o paths) son secuencias de nodos unidos por flechas o aristas.



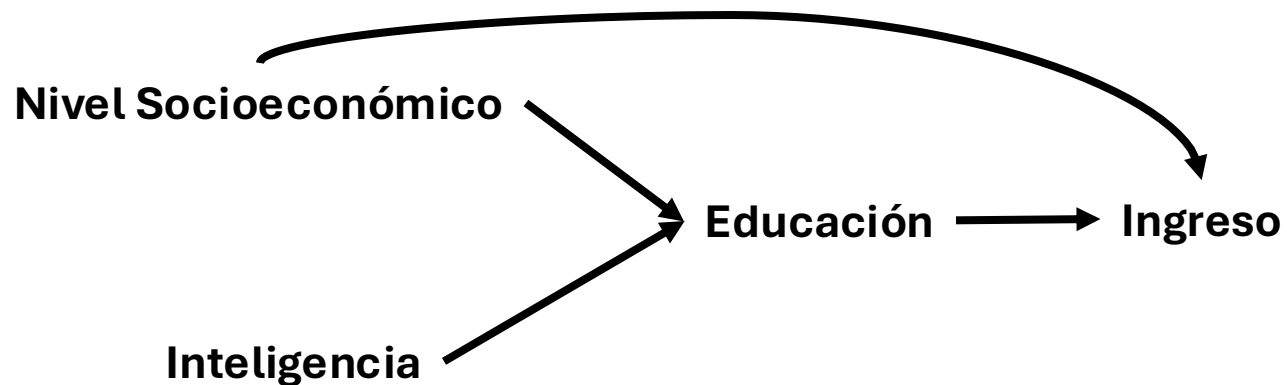
Los DAGs presentan las siguientes características:

- Dirigidos: Las flechas tienen una sola dirección
- Acíclicos: Ningún camino puede terminar donde empezó
 - Por lo general se entiende la cronología de las variables de derecha a izquierda

Un diagrama para establecer nuestros supuestos

Dibujar un DAG

- Tanto la presencia como ausencia de flechas implican un supuesto sobre las relaciones causales
- El expresar estos supuestos en formato de DAG no los hace más válidos, pero sí los hace explícitos, facilita debatir acerca de ellos, y nos hace ser conscientes de que nuestro análisis parte sobre la base de tales supuestos

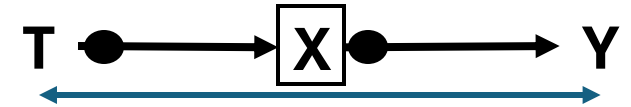
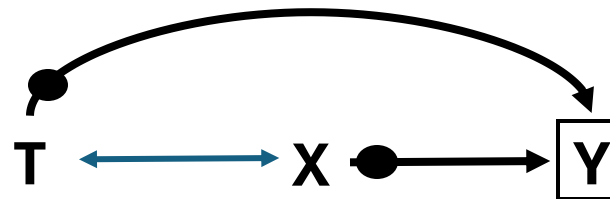
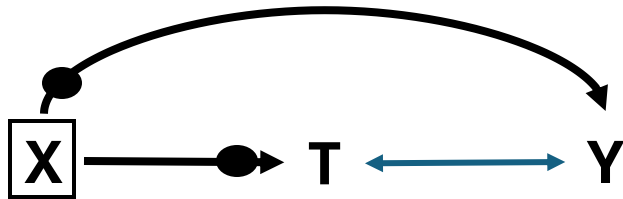


- Una flecha ausente es un supuesto más fuerte que el de una presente
- Deben ser justificados mediante el marco teórico y la evidencia empírica disponible

Estructuras de diagramas causales

Existen 3 tipos de estructuras básicas en los gráficos causales

- Bifurcaciones o Forks
- Confluencias o Colliders
- Cadenas causales



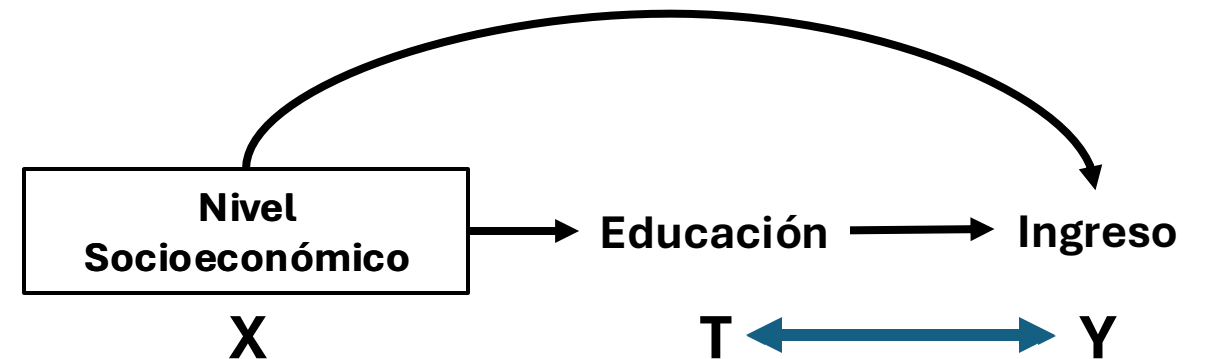
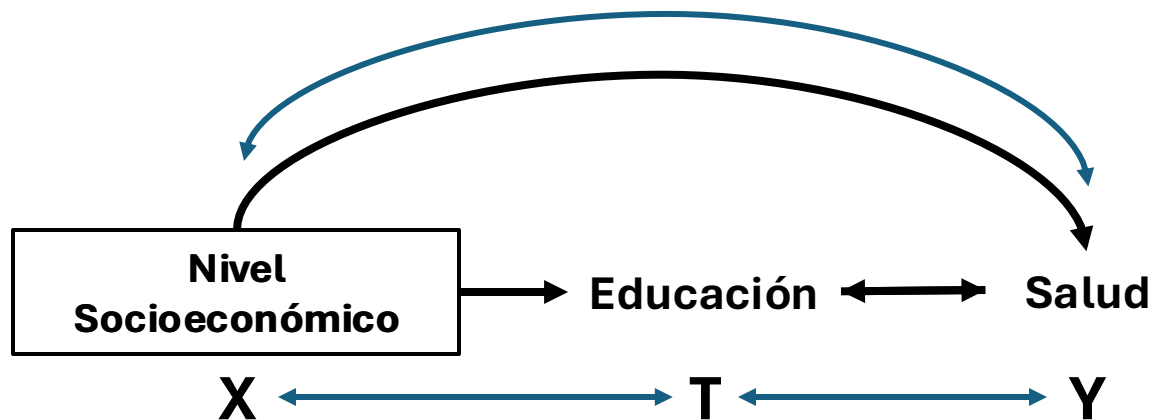
Los modelos causales son a su vez modelos estadísticos

- Los efectos causales implican asociaciones (correlaciones)
- La ausencia de efectos causales implica independencia estadística
- Los sesgos sistemáticos pueden ser identificados a través de estas asociaciones

Estructuras de diagramas causales

Bifurcaciones o forks

- La variable **X** tiene un efecto sobre **T** y sobre **Y** simultáneamente
- Si controlamos por **X**, entonces el flujo de asociación se detendrá
- Evitaremos el sesgo provocado por **X**
- Estas variables se suelen conocer como confounding variables o puertas traseras
- El sesgo por confounding es el primer tipo de sesgo sistemático

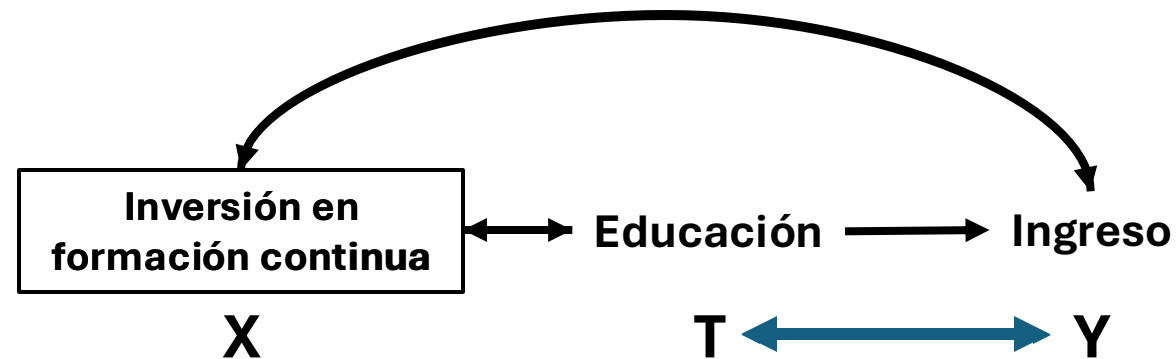


Estructuras de diagramas causales



Confluencias o Colliders

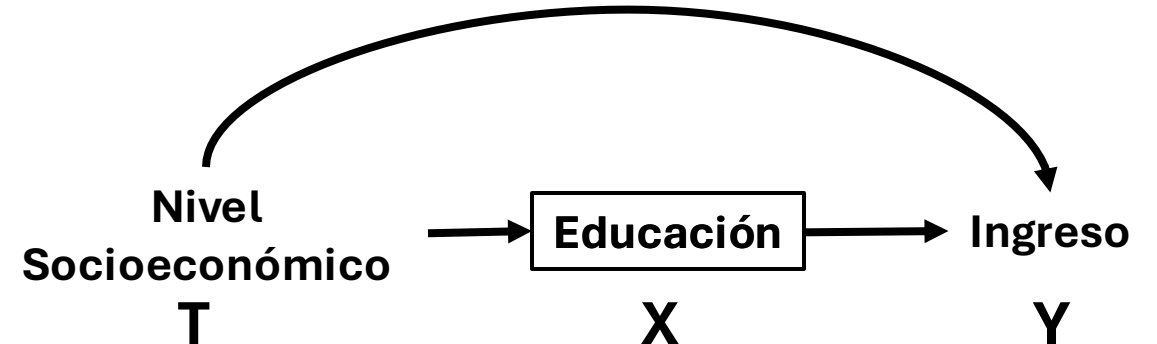
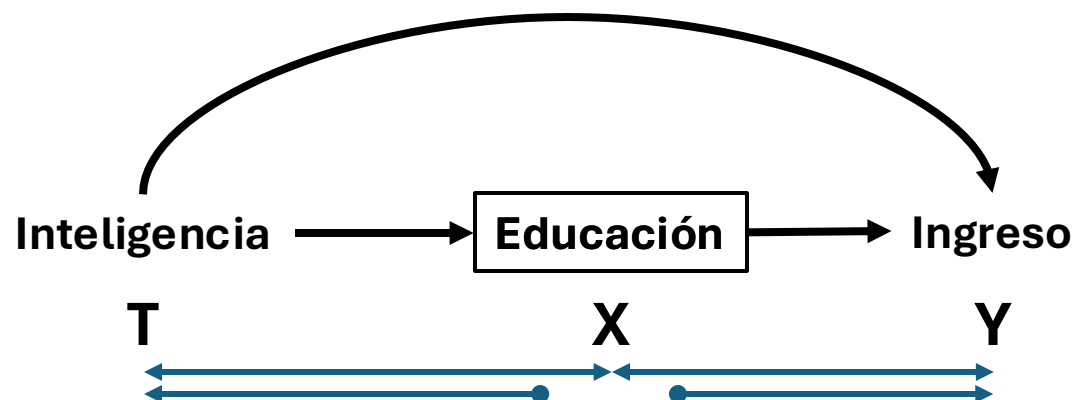
- Las variables **T** y **Y** tienen un efecto sobre **X**
- Cuando controlamos por un efecto común entre el tratamiento y el resultado se provoca un sesgo.
- La segunda fuente significativa de sesgo es lo que llamaremos sesgo de selección.



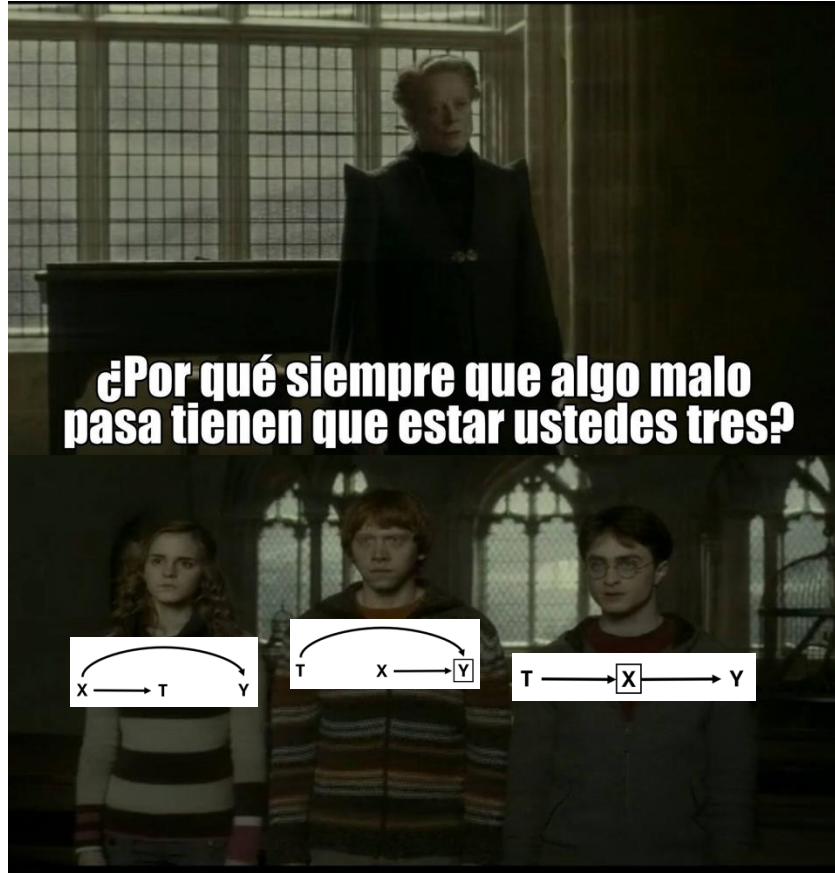
Estructuras de diagramas causales

Cadenas causales

- En una cadena causal, el efecto de **T** sobre **Y** ocurre sólo a través de **X**
 - Si controlamos por **T**, entonces el flujo de asociación se detendrá.
-
- El sesgo de selección también ocurre cuando condicionamos en un mediador entre la variable de tratamiento y la variable resultado.



D-separación



Reglas de D-separación

- Dos variables están D-separadas si todos los caminos entre ellas están bloqueados.
- D-separación = independencia estadística (no asociación)

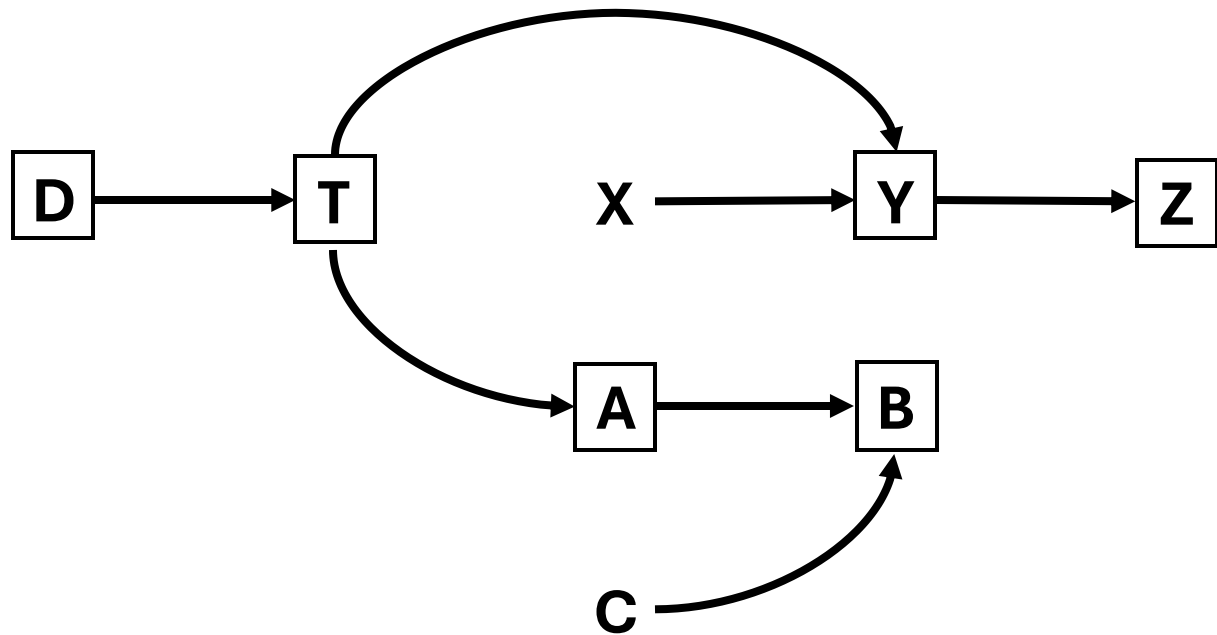
Existen 4 reglas de D-separación pero estas pueden ser resumidas en las siguientes dos condiciones.

- Un camino está bloqueado si y solo si:
 - Contiene un colisionador que no ha sido condicionado y que no tiene descendientes que hayan sido condicionados.
 - Contiene un no colisionador que ha sido condicionado.

D-separación

Un camino está bloqueado si y solo si:

- Contiene un colisionador que no ha sido condicionado y que no tiene descendientes que hayan sido condicionados.
- Contiene un no colisionador que ha sido condicionado.



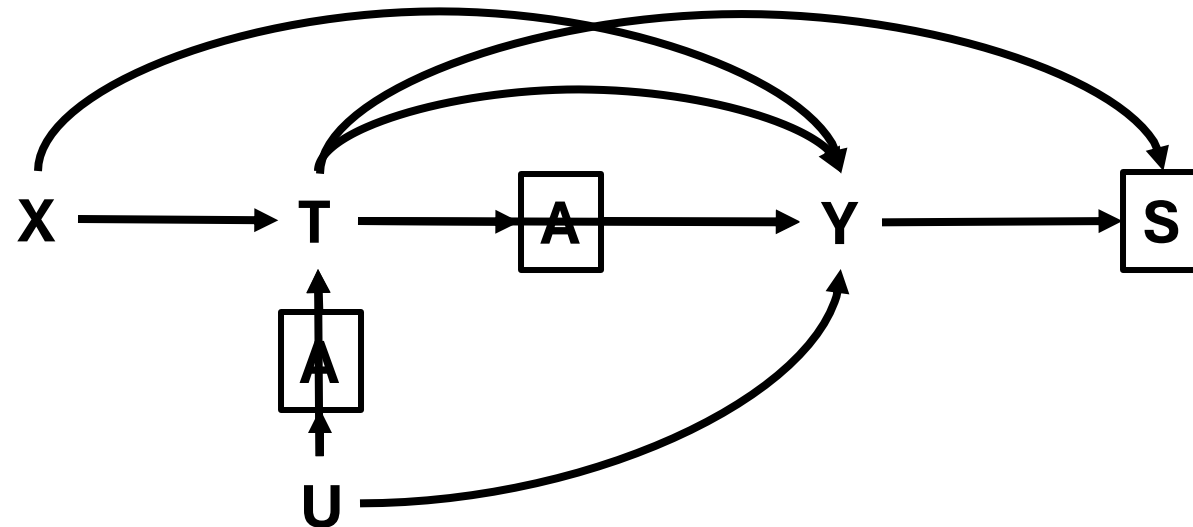
- Dos variables son independientes incondicionalmente si están D-separadas sin ninguna variable en el camino condicionada.
- Dos variables son independientes condicionalmente si están D-separadas con alguna variable en el camino condicionada.

¿Cómo dibujar un diagrama causal?

Un diagrama causal no pretende describir la realidad lo más precisa posible

¿Qué debe incluir un diagrama causal?

- Las variables confounding que pudieran generar sesgo (medibles y no medibles)
- Las variables colliders que pudieran estar siendo condicionadas
- Los mediadores que quisiéramos condicionar para bloquear algún camino



Métodos econométricos como diagramas causales

Toda estrategia de inferencia causal tiene su diagrama asociado

- Variables instrumentales
- Efectos fijos y variables
- Propensity Score

